

TITLE:

Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci

AUTHOR LIST & AFFILIATIONS:

Fredrick R. Schumacher^{1,2,142}, Ali Amin Al Olama^{3,4,142}, Sonja I. Berndt^{5,142}, Sara Benlloch^{3,6}, Mahbub Ahmed⁶, Edward J. Saunders⁶, Tokhir Dadaev⁶, Daniel Leongamornlert⁶, Ezequiel Anokian⁶, Clara Cieza-Borrella⁶, Chee Goh⁶, Mark N. Brook⁶, Xin Sheng⁷, Laura Fachal^{8,9}, Joe Dennis³, Jonathan Tyrer³, Kenneth Muir^{10,11}, Artitaya Lophatananon^{10,11}, Victoria L. Stevens¹², Susan M. Gapstur¹², Brian D. Carter¹², Catherine M. Tangen¹³, Phyllis J. Goodman¹³, Ian M. Thompson Jr.¹⁴, Jyotsna Batra^{15,16}, Suzanne Chambers^{17,18}, Leire Moya^{15,16}, Judith Clements^{15,16}, Lisa Horvath^{19,20}, Wayne Tilley²¹, Gail P. Risbridger^{22,23}, Henrik Gronberg²⁴, Markus Aly^{25,26}, Tobias Nordström^{24,27}, Paul Pharoah^{3,9}, Nora Pashayan^{9,28}, Johanna Schleutker^{29,30}, Teuvo L.J. Tammela³¹, Csilla Sipeky²⁹, Anssi Auvinen³², Demetrius Albanes⁵, Stephanie Weinstein⁵, Alicia Wolk^{33,34}, Niclas Håkansson³³, Catharine M.L. West³⁵, Alison M. Dunning⁹, Neil Burnet³⁶, Lorelei A. Mucci³⁷, Edward Giovannucci³⁷, Gerald L. Andriole³⁸, Olivier Cussenot^{39,40}, Géraldine Cancel-Tassin^{39,40}, Stella Koutros⁵, Laura E. Beane Freeman⁵, Karina Dalsgaard Sorensen^{41,42}, Torben Falck Orntoft^{41,42}, Michael Borre^{42,43}, Lovise Maehle⁴⁴, Eli Marie Grindedal⁴⁴, David E. Neal⁴⁵⁻⁴⁷, Jenny L. Donovan⁴⁸, Freddie C. Hamdy⁴⁷, Richard M. Martin⁴⁸, Ruth C. Travis⁴⁹, Tim J. Key⁴⁹, Robert J. Hamilton⁵⁰, Neil E. Fleshner⁵⁰, Antonio Finelli⁵¹, Sue Ann Ingles⁷, Mariana C. Stern⁷, Barry S. Rosenstein^{52,53}, Sarah L. Kerns⁵⁴, Harry Ostrer⁵⁵, Yong-Jie Lu⁵⁶, Hong-Wei Zhang⁵⁷, Ninghan Feng⁵⁸, Xueying Mao⁵⁶, Xin Guo⁵⁹, Guomin Wang⁶⁰, Zan Sun⁶¹, Graham G. Giles^{62,63}, Melissa C. Southey⁶⁴, Robert J. MacInnis^{62,63}, Liesel M. FitzGerald^{62,65}, Adam S. Kibel⁶⁶, Bettina F. Drake³⁸, Ana Vega⁶⁷, Antonio Gómez-Caamaño⁶⁸, Robert Szulkin^{69,70}, Martin Eklund²⁴, Manolis Kogevinas⁷¹⁻⁷⁴, Javier Llorca^{72,75}, Gemma Castaño-Vinyals⁷¹⁻⁷⁴, Kathryn L. Penney⁷⁶, Meir Stampfer⁷⁶, Jong Y. Park⁷⁷, Thomas A. Sellers⁷⁷, Hui-Yi Lin⁷⁸, Janet L. Stanford^{79,80}, Cezary Cybulski⁸¹, Dominika Wokolorczyk⁸¹, Jan Lubinski⁸¹, Elaine A. Ostrander⁸², Milan S. Geybels⁷⁹, Børge G. Nordestgaard^{83,84}, Sune F. Nielsen^{83,84}, Maren Weischer⁸⁴, Rasmus Bisbjerg⁸⁵, Martin Andreas Røder⁸⁶, Peter Iversen⁸⁶, Hermann Brenner⁸⁷⁻⁸⁹, Katarina Cuk⁸⁷, Bernd Holleczek⁹⁰, Christiane Maier⁹¹, Manuel Luedeke⁹¹, Thomas Schnoeller⁹², Jeri Kim⁹³, Christopher J. Logothetis⁹³, Esther M. John^{94,95}, Manuel R. Teixeira^{96,97}, Paula Paulo⁹⁶, Marta Cardoso⁹⁶, Susan L. Neuhausen⁹⁸, Linda Steele⁹⁸, Yuan Chun Ding⁹⁸, Kim De Ruyck⁹⁹, Gert De Meerleer⁹⁹, Piet Ost¹⁰⁰, Azad Razack¹⁰¹, Jasmine Lim¹⁰¹, Soo-Hwang Teo¹⁰², Daniel W. Lin^{79,103}, Lisa F. Newcomb^{79,103}, Davor Lessel¹⁰⁴, Marija Gamulin¹⁰⁵, Tomislav Kulis¹⁰⁶, Radka Kaneva¹⁰⁷, Nawaid Usmani^{108,109}, Sandeep Singhal^{108,109}, Chavdar Slavov¹¹⁰, Vanio Mitev¹⁰⁷, Matthew Parliament^{108,109}, Frank Claessens¹¹¹, Steven Joniau¹¹², Thomas Van den Broeck^{111,112}, Samantha Larkin¹¹³, Paul A. Townsend¹¹⁴, Claire Aukim-Hastie¹¹⁵, Manuela Gago Dominguez^{116,117}, Jose Esteban Castelao¹¹⁸, Maria Elena Martinez¹¹⁹, Monique J. Roobol¹²⁰, Guido Jenster¹²⁰, Ron H.N. van Schaik¹²¹, Florence Menegaux¹²², Thérèse Truong¹²², Yves Akoli Koudou¹²², The Profile Study Steering Committee¹²³, Jianfeng Xu¹²⁴, Kay-Tee Khaw¹²⁵, Lisa Cannon-Albright^{126,127}, Hardev Pandha¹¹⁵, Agnieszka Michael¹¹⁵, Stephen N. Thibodeau¹²⁸, Shannon K. McDonnell¹²⁹, Daniel J. Schaid¹²⁹, Sara Lindstrom¹³⁰, Constance Turman¹³¹, Jing Ma⁷⁶, David J. Hunter¹³¹, Elio Riboli¹³², Afshan Siddiq¹³³, Federico Canzian¹³⁴, Laurence N. Kolonel¹³⁵, Loic Le Marchand¹³⁵, Robert N. Hoover⁵, Mitchell J. Machiela⁵, Zuxi Cui¹, Peter Kraft¹³¹, Australian Prostate Cancer BioResource¹³⁶, The IMPACT Study Steering Committee¹³⁶, Canary PASS Investigators¹³⁶, BPC3 (Breast and Prostate Cancer Cohort

Consortium)¹³⁶, The PRACTICAL (Prostate Cancer Association Group to Investigate Cancer-Associated Alterations in the Genome) Consortium¹³⁶, CAPS (Cancer of the Prostate in Sweden)¹³⁶, PEGASUS (Prostate Cancer Genome-wide Association Study of Uncommon Susceptibility Loci)¹³⁶, The GAME-ON/ELLIPSE Consortium¹³⁶, Christopher I. Amos¹³⁷⁻¹³⁹, David V. Conti^{7,143}, Douglas F. Easton^{3,9,143}, Fredrik Wiklund^{24,143}, Stephen J. Chanock^{5,143}, Brian E. Henderson^{7,141,143}, ZSofia Kote-Jarai^{6,143}, Christopher A. Haiman^{7,143}, Rosalind A. Eeles^{6,140,143}

- 1 Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, Ohio, USA
- 2 Seidman Cancer Center, University Hospitals, Cleveland, Ohio, USA
- 3 Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Strangeways Research Laboratory, Cambridge, UK.
- 4 University of Cambridge, Department of Clinical Neurosciences, Cambridge, UK
- 5 Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Bethesda, Maryland, USA
- 6 The Institute of Cancer Research, London, UK
- 7 Department of Preventive Medicine, Keck School of Medicine, University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, California, USA
- 8 Fundación Pública Galega de Medicina Xenómica-SERGAS, Grupo de Medicina Xenómica, CIBERER, IDIS, Santiago de Compostela, Spain
- 9 Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Strangeways Laboratory, Cambridge, UK
- 10 Division of Population Health, Health Services Research and Primary Care, University of Manchester, Manchester, UK
- 11 Warwick Medical School, University of Warwick, Coventry, UK
- 12 Epidemiology Research Program, American Cancer Society, Atlanta, Georgia, USA
- 13 SWOG Statistical Center, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA
- 14 CHRISTUS Santa Rosa Hospital - Medical Center, San Antonio, Texas, USA
- 15 Institute of Health and Biomedical Innovation and School of Biomedical Sciences, Queensland University of Technology, Brisbane, Queensland, Australia
- 16 Australian Prostate Cancer Research Centre-Qld, Translational Research Institute, Brisbane, Queensland, Australia
- 17 Menzies Health Institute of Queensland, Griffith University, Queensland, Australia
- 18 Cancer Council Queensland, Fortitude Valley, QLD, Australia
- 19 Chris O'Brien Lifehouse (COBLH), Camperdown, NSW, Australia
- 20 Garvan Institute of Medical Research, Sydney, NSW, Australia
- 21 Dame Roma Mitchell Cancer Research Centre, University of Adelaide, Adelaide, South Australia, Australia
- 22 Monash Biomedicine Discovery Institute Cancer Program, Prostate Cancer Research Program, Department of Anatomy and Developmental Biology, Monash University, Victoria, Australia

- 23 Cancer Research Division, Peter MacCallum Cancer Centre, Melbourne, Australia
- 24 Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden
- 25 Department of Urology, Karolinska University Hospital, Stockholm, Sweden
- 26 Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden
- 27 Department of Clinical Sciences at Danderyds Hospital, Karolinska Institutet, Stockholm, Sweden
- 28 Department of Applied Health Research, University College London, London, UK
- 29 Institute of Biomedicine, University of Turku, Finland
- 30 Tyks Microbiology and Genetics, Department of Medical Genetics, Turku University Hospital, Finland
- 31 Department of Urology, Tampere University Hospital and Faculty of Medicine and Life Sciences, University of Tampere, Finland
- 32 Department of Epidemiology, School of Health Sciences, University of Tampere, Tampere, Finland
- 33 Division of Nutritional Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, Sweden
- 34 Department of Surgical Sciences, Uppsala University, Uppsala, Sweden
- 35 Division of Cancer Sciences, University of Manchester, Manchester Cancer Research Centre, Manchester Academic Health Science Centre, and The Christie NHS Foundation Trust, Manchester, UK
- 36 University of Cambridge, Department of Oncology, Oncology Centre, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK
- 37 Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA
- 38 Washington University School of Medicine, St. Louis, Missouri, USA
- 39 UPMC Sorbonne Universites, GRC N°5 ONCOTYPE-URO, Tenon Hospital, Paris, France
- 40 CeRePP, Tenon Hospital, Paris, France
- 41 Department of Molecular Medicine, Aarhus University Hospital, Denmark
- 42 Department of Clinical Medicine, Aarhus University, Denmark
- 43 Department of Urology, Aarhus University Hospital, Denmark
- 44 Department of Medical Genetics, Oslo University Hospital, Oslo, Norway
- 45 University of Cambridge, Department of Oncology, Addenbrooke's Hospital, Cambridge, UK
- 46 Cancer Research UK Cambridge Research Institute, Li Ka Shing Centre, Cambridge, UK
- 47 Nuffield Department of Surgical Sciences, University of Oxford, Oxford, UK, Faculty of Medical Science, University of Oxford, John Radcliffe Hospital, Oxford, UK
- 48 School of Social and Community Medicine, University of Bristol, Bristol, UK
- 49 Cancer Epidemiology Unit, Nuffield Department of Population Health, University of Oxford, Oxford, UK
- 50 Department of Surgical Oncology, Princess Margaret Cancer Centre, Toronto, Canada

- 51 Division of Urology, Princess Margaret Cancer Centre, Toronto, Canada
- 52 Department of Radiation Oncology, Icahn School of Medicine at Mount Sinai,
New York, New York, USA
- 53 Department of Genetics and Genomic Sciences, Icahn School of Medicine at
Mount Sinai, New York, New York, USA
- 54 Department of Radiation Oncology, University of Rochester Medical Center,
Rochester, New York, USA
- 55 Professor of Pathology and Pediatrics, Albert Einstein College of Medicine,
Bronx, New York, USA
- 56 Centre for Molecular Oncology, Barts Cancer Institute, Queen Mary University of
London, John Vane Science Centre, London, UK
- 57 Second Military Medical University, Shanghai, China
- 58 Wuxi Second Hospital, Nanjing Medical University, Wuxi, Jiangzhu Province,
China
- 59 Department of Urology, The First Affiliated Hospital, Chongqing Medical
University, Chongqing, China
- 60 Department of Urology, Zhongshan Hospital, Fudan University Medical College,
Shanghai, China
- 61 The People's Hospital of Liaoning Province, The People's Hospital of China
Medical University, Shenyang, China
- 62 Cancer Epidemiology and Intelligence Division, Cancer Council Victoria,
Melbourne, Victoria, Australia
- 63 Centre for Epidemiology and Biostatistics, Melbourne School of Population and
Global Health, The University of Melbourne, Melbourne, Australia
- 64 Precision Medicine, School and Clinical Sciences at Monash Health, Monash
University, Clayton, Victoria, Australia
- 65 Menzies Institute for Medical Research, University of Tasmania, Hobart,
Australia
- 66 Division of Urologic Surgery, Brigham and Womens Hospital, Boston,
Massachusetts, USA
- 67 Fundación Pública Galega de Medicina Xenómica-SERGAS, Grupo de Medicina
Xenómica, CIBERER, IDIS, Santiago de Compostela, Spain
- 68 Department of Radiation Oncology, Complejo Hospitalario Universitario de
Santiago, SERGAS, Santiago de Compostela, Spain
- 69 Division of Family Medicine, Department of Neurobiology, Care Science and
Society, Karolinska, Institutet, Huddinge, Sweden
- 70 Scandinavian Development Services, Danderyd, Sweden
- 71 ISGlobal, Centre for Research in Environmental Epidemiology (CREAL),
Barcelona, Spain
- 72 CIBER Epidemiología y Salud Pública (CIBERESP), Madrid, Spain
- 73 IMIM (Hospital del Mar Research Institute), Barcelona, Spain
- 74 Universitat Pompeu Fabra (UPF), Barcelona, Spain
- 75 University of Cantabria-IDIVAL, Santander, Spain
- 76 Channing Division of Network Medicine, Department of Medicine, Brigham and
Women's Hospital/Harvard Medical School, Boston, Massachusetts, USA
- 77 Department of Cancer Epidemiology, Moffitt Cancer Center, Tampa, Florida, USA

- 78 Biostatistics Program, School of Public Health, Louisiana State University Health Sciences Center, New Orleans, Louisiana, USA
- 79 Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA
- 80 Department of Epidemiology, School of Public Health, University of Washington, Seattle, Washington, USA
- 81 International Hereditary Cancer Center, Department of Genetics and Pathology, Pomeranian Medical University, Szczecin, Poland
- 82 National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, USA
- 83 Faculty of Health and Medical Sciences, University of Copenhagen, Denmark
- 84 Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark
- 85 Department of Urology, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark
- 86 Copenhagen Prostate Cancer Center, Department of Urology, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark
- 87 Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany
- 88 German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany
- 89 Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany
- 90 Saarland Cancer Registry, Saarbrücken, Germany
- 91 Institute for Human Genetics, University Hospital Ulm, Ulm, Germany
- 92 Department of Urology, University Hospital Ulm, Germany
- 93 The University of Texas - MD Anderson Cancer Center, Department of Genitourinary Medical Oncology, Houston, Texas, USA
- 94 Cancer Prevention Institute of California, Fremont, California, USA
- 95 Department of Health Research & Policy (Epidemiology) and Stanford Cancer Institute, Stanford University School of Medicine, Stanford, California, USA
- 96 Department of Genetics, Portuguese Oncology Institute of Porto, Porto, Portugal
- 97 Biomedical Sciences Institute (ICBAS), University of Porto, Porto, Portugal
- 98 Department of Population Sciences, Beckman Research Institute of the City of Hope, Duarte, California, USA
- 99 Ghent University, Faculty of Medicine and Health Sciences, Basic Medical Sciences, Ghent, Belgium
- 100 Ghent University Hospital, Department of Radiotherapy, Ghent, Belgium
- 101 Department of Surgery, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia
- 102 Cancer Research Malaysia (CRM), Outpatient Centre, Subang Jaya Medical Centre, Selangor, Malaysia
- 103 Department of Urology, University of Washington, Seattle, Washington, USA
- 104 Institute of Human Genetics, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

- 105 Division of Medical Oncology, Urogenital Unit, Department of Oncology,
University Hospital Centre Zagreb, Zagreb, Croatia
- 106 Department of Urology, University Hospital Center Zagreb, University of Zagreb
School of Medicine, Zagreb, Croatia
- 107 Molecular Medicine Center, Department of Medical Chemistry and Biochemistry,
Medical University of Sofia, Sofia, Bulgaria
- 108 Department of Oncology, Cross Cancer Institute, University of Alberta,
Edmonton, Alberta, Canada
- 109 Division of Radiation Oncology, Cross Cancer Institute, Edmonton, Alberta,
Canada
- 110 Department of Urology and Alexandrovska University Hospital, Medical
University of Sofia, Sofia, Bulgaria
- 111 Molecular Endocrinology Laboratory, Department of Cellular and Molecular
Medicine, KU Leuven, Leuven, Belgium
- 112 Department of Urology, University Hospitals Leuven, Leuven, Belgium
- 113 The University of Southampton, Southampton General Hospital, Southampton,
UK
- 114 Division of Cancer Sciences, Manchester Cancer Research Centre, Faculty of
Biology, Medicine and Health, Manchester Academic Health Science Center, NIHR
Manchester Biomedical Research Centre, Health Innovation Manchester,
University of Manchester, Manchester, UK
- 115 The University of Surrey, Guildford, Surrey, UK
- 116 Genomic Medicine Group, Galician Foundation Genomic Medicine, Instituto de
Investigacion Sanitaria de Santiago de Compostela (IDIS), Complejo Hospitalario
Universitario de Santiago, Servicio Galego de Saúde, SERGAS, Santiago De
Compostela, Spain
- 117 University of California San Diego, Moores Cancer Center, La Jolla, California,
USA
- 118 Genetic Oncology Unit, CHUVI Hospital, Complejo Hospitalario Universitario de
Vigo, Instituto de Investigación Biomédica Galicia Sur (IISGS), Vigo SERGAS,
Spain
- 119 University of California San Diego, Moores Cancer Center, Department of Family
Medicine and Public Health, University of California San Diego, La Jolla,
California, USA
- 120 Department of Urology, Erasmus University Medical Center, Rotterdam, the
Netherlands
- 121 Department of Clinical Chemistry, Erasmus University Medical Center,
Rotterdam, the Netherlands
- 122 Cancer & Environment Group, Center for Research in Epidemiology and
Population Health (CESP), INSERM, University Paris-Sud, University Paris-
Saclay, Villejuif, France
- 123 [http://www.cancerresearchuk.org/about-cancer/find-a-clinical-trial/a-study-
find-out-looking-gene-changes-would-be-useful-in-screening-for-prostate-
cancer-profile-pilot](http://www.cancerresearchuk.org/about-cancer/find-a-clinical-trial/a-study-find-out-looking-gene-changes-would-be-useful-in-screening-for-prostate-cancer-profile-pilot)
- 124 Program for Personalized Cancer Care, NorthShore University HealthSystem,
Evanston, Illinois, USA

- 125 Clinical Gerontology Unit, University of Cambridge, Cambridge, UK
- 126 Division of Genetic Epidemiology, Department of Medicine, University of Utah
School of Medicine, Salt Lake City, Utah, USA
- 127 George E. Wahlen Department of Veterans Affairs Medical Center, Salt Lake City,
Utah, USA
- 128 Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester,
Minnesota, USA
- 129 Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester,
Minnesota, USA
- 130 Department of Epidemiology, Health Sciences Building, University of
Washington, Seattle, Washington, USA
- 131 Program in Genetic Epidemiology and Statistical Genetics, Department of
Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts,
USA
- 132 Department of Epidemiology and Biostatistics, School of Public Health, Imperial
College London, London, UK
- 133 Genomics England, Queen Mary University of London, Dawson Hall,
Charterhouse Square, London, UK
- 134 Genomic Epidemiology Group, German Cancer Research Center, (DKFZ),
Heidelberg, Germany
- 135 Epidemiology Program, University of Hawaii Cancer Center, Honolulu, Hawaii,
USA
- 136 A list of members and affiliations appears in the Supplementary Note
- 137 Department of Biomedical Data Science, Geisel School of Medicine at Dartmouth,
Lebanon, New Hampshire, USA
- 138 Department of Molecular and Systems Biology, Geisel School of Medicine at
Dartmouth, Hanover, New Hampshire, USA
- 139 Norris Cotton Cancer Center, Geisel School of Medicine at Dartmouth, Lebanon,
New Hampshire, USA
- 140 Royal Marsden NHS Foundation Trust, London, UK
- 141 In memoriam
- 142 These authors contributed equally to this work
- 143 These authors jointly directed this work

Correspondence should be addressed to F.R.S. (frs2@case.edu), A. Al Olama
(aa461@medschl.cam.ac.uk) and R.A.E. (ros.eeles@icr.ac.uk)

ABSTRACT:

Currently genome-wide association studies (GWAS) and fine-mapping efforts have identified over 100 prostate cancer (PrCa) susceptibility loci. We meta-analyzed genotype data from a custom high-density array of 46,939 PrCa cases and 27,910 controls of European ancestry with previously genotyped data of 32,255 PrCa cases and 33,202 controls of European ancestry. Our analysis identified 62 novel loci associated ($P < 5.0 \times 10^{-8}$) with PrCa, and a locus significantly associated with early-onset PrCa (≤ 55 years). Our findings include missense variants rs1800057 (OR=1.16; $P=8.2 \times 10^{-9}$; G>C [Pro1054Arg]) in *ATM* and rs2066827 (OR=1.06; $P=2.3 \times 10^{-9}$; T>G [Val109Gly]) in *CDKN1B*. The combination of all loci captures 28.4% of the PrCa familial relative risk and a polygenic risk score confers an elevated PrCa risk for men in the 90-99%-ile (RR=2.69; 95%CI: 2.55-2.82) and 1%-ile (RR=5.71; 95%CI: 5.04-6.48) risk stratum compared to the population average. These findings improve risk prediction, enhance fine-mapping, and provide insight into the underlying biology of PrCa¹.

MAIN TEXT:

Although prostate cancer (PrCa) is the most common non-cutaneous cancer among men in the Western world and 1 in 7 men will be diagnosed during their lifetime², very few modifiable risk factors have been established³. Epidemiological studies have identified age, a positive family history and race/ethnicity as the most prominent risk factors for PrCa⁴⁻⁷. PrCa incidence is highest among men of African ancestry, followed by men of European and Asian ancestry. These ancestral differences of PrCa risk, in conjunction with studies demonstrating the influence of family history^{8,9}, highlight the contribution of genetics in PrCa etiology¹⁰. Our previous work, utilizing a multiplicative model, estimated over 1,800 common single nucleotide polymorphisms (SNPs) independently contribute to PrCa risk among populations of European ancestry¹¹. Genome-wide association studies (GWAS) have reported over 100 of these PrCa variants across multi-ethnic populations, with the vast majority being identified in populations of European ancestry¹²⁻²⁹.

To facilitate additional novel discovery of PrCa genetic risk factors we developed a custom high-density genotyping array, the OncoArray, including a 260K SNP backbone designed to adequately tag most common genetic variants (MAF>5% in Europeans),

and 310K SNPs from the meta-analyses of five cancers (breast, colorectal, lung, ovarian, and prostate)³⁰. Approximately 80,000 PrCa-specific markers derived from our previous multi-ethnic meta-analysis¹² (including populations of European, African American, Japanese, and Latino ancestry), fine-mapping of known PrCa loci, and candidate SNPs nominated by study collaborators were included on the OncoArray. We assembled a new PrCa sample series from 52 studies to genotype with the OncoArray (**Supplementary Tables 1 & 2**). After applying rigorous quality control criteria and removing overlapping samples from previous studies, our OncoArray sample yielded 46,939 PrCa cases and 27,910 controls without a known diagnosis of PrCa of European ancestry for analysis (see **Online Methods, Supplementary Table 3**). Genotypes were phased and imputed to the cosmopolitan panel of the 1000 Genomes Project (1KGP; 2014 June release) using SHAPEIT³¹ and IMPUTEv2³² software (**Online Methods, Supplementary Table 3**). We performed a fixed-effects meta-analysis combining the summary statistics from our OncoArray analysis, and seven previous PrCa GWAS or high-density SNP panels of European ancestry imputed to 1KGP. The final meta-analysis included 79,194 PrCa cases and 61,112 controls without a known diagnosis of PrCa (**Figure 1**).

Study- and consortia-specific meta-analyses were performed to identify novel PrCa loci. We established a P -value threshold of 5.0×10^{-8} to determine genome-wide significance. Our large sample size enabled several stratified meta-analyses focusing on key clinical and biological parameters (**Online Methods, Supplementary Tables 4 & 5**). All analyses used a likelihood ratio test to minimize bias from rare variants and a logistic regression framework was used for all analyses, except for Gleason score where linear regression was utilized. The genotype dosages were incorporated in an allelic genetic model. The average λ_{1000} , an inflation statistic calibrated to a sample size of 1000 cases and 1000 controls³³, across the eight GWAS studies was 1.02 (range: 0.98-1.09) and 1.00 for the overall meta-analysis (**Supplementary Table 6**). Our novel findings excluded variants within defined fine-mapped regions of previously reported PrCa loci (**Supplementary Table 7**).

After the exclusion of all known susceptibility regions (fine-mapping coordinates provided in **Supplementary Table 7 & Supplementary Note**) we identified 64 loci associated with overall PrCa susceptibility and one with early-onset ($P < 5.0 \times 10^{-8}$) in the meta-

analysis (**Supplementary Figure 1**), where 53 were imputed and 12 were genotyped using the OncoArray . The cluster plots for the genotyped makers are presented in **Supplementary Figure 2**. Although a majority of the imputed markers were of high quality with an average imputed r^2 greater than 0.80 for 61 of the 65 loci across all contributing GWAS (**Supplementary Table 8**), we closely examined four variants with a poor imputation quality score ($r^2 < 0.80$) in the OncoArray samples by inspecting linkage disequilibrium (LD) plots including only genotyped SNPs from the OncoArray and performing an imputation quality control assessment (**Online Methods**). After reviewing the LD plots and the imputation QC, we determined loci rs6602880 and rs144166867 are likely false positives due to imputation artifacts (**Supplementary Figure 3; Supplementary Table 9**). Overall, we identified 62 novel loci associated with overall PrCa risk and one novel loci associated with early-onset (**Table 1**). The consortia specific associations were consistent across the eight contributing GWAS studies (**Supplementary Table 10**).

We performed several stratified analyses defined by clinical and population parameters. We detected a novel variant, rs138004030, significantly associated with early-onset disease (**Table 1**), but only nominally significant for overall PrCa risk ($P=0.02$). In addition, we detected four markers significantly associated ($P < 5 \times 10^{-8}$) with advanced PrCa and two markers associated with early-onset PrCa (**Supplementary Table 11**). However, the case-only analyses of these markers indicated marginal statistical significance ($P < 1.0 \times 10^{-3}$). Additionally, these markers were in LD with nearby index markers associated with overall PrCa and not significantly associated with overall aggressive disease after adjusting for the index marker (**Supplementary Table 11**). A similar association pattern was observed for rs111599055, which was in LD with marker rs7295014 ($r^2=0.54$) associated with overall disease. The early-onset marker rs77777548 is independent of novel and known PrCa loci. However, the marker is relatively rare (EAF<0.02), indicated as monomorphic in 1KGP, and has a moderate imputation quality score (average $r^2=0.57$) hence we did not include it in further analyses.

Among the 63 novel associations, 38 variants are located within gene-rich regions (**Supplementary Table 12**): intronic (32 SNPs), missense (4 SNPs), and 3'-UTR (2 SNPs). eQTL analyses of the TCGA database identified statistically significant

associations ($P < 0.05$; **Supplementary Table 12**) in normal PrCa tissue for 17 of the novel associations, including both 3'UTR SNPs and 11 of the 32 intronic SNPs. *Cis* eQTL associations were identified for 3'UTR variant rs1048169 with *HAUS6* (3'UTR) and intronic variants rs182314334 with *MBNL1*, rs4976790 with *COL23A1*, rs9469899 with *UHRF1BP1*, rs878987 with *B3GAT1*, rs11629412 with *PAX9*, and rs11666569 with *MYO9B*. The eQTL associations are consistent with the observed PrCa-SNP associations, given we assessed colocalization between the GWAS and eQTL SNPs. The TCGA data analysis failed to identify an eQTL association with any of the four missense SNPs.

We assessed the association of our newly discovered loci with prostate-specific antigen (PSA) levels using a series of disease-free controls (N=9,090; see **Online Methods**). Among the 48 available loci we observed a significant association for rs8093601 ($P = 5.0 \times 10^{-4}$; **Supplementary Table 13**) after correcting for multiple testing ($P = 0.05/48 = 1.0 \times 10^{-3}$). This marker lies near *MBD2*, methyl-CpG binding domain protein 2, and has not been previously associated with either PrCa risk or PSA levels. The effect estimates of PrCa clinical features and overall PrCa did not differ (**Supplementary Table 14**). LD plots incorporating several functional annotation features for each of the 63 novel markers is presented in **Supplementary Figure 4**.

Several strong candidate genes were identified among the PrCa susceptibility loci, including *ATM*, a key gene within the DNA damage response pathway, in which truncating variants contribute towards PrCa susceptibility and progression, particularly aggressive PrCa^{34,35}. The index variant within this region is the missense variant rs1800057, exerting a modest increased risk of PrCa (OR=1.16; $P = 8.15 \times 10^{-9}$; G>C [Pro1054Arg]; **Figure 2, Panel A**). Although rs1800057 is designated 'benign' by ClinVar (<http://www.ncbi.nlm.nih.gov/clinvar/>), it was previously suggested to be associated with a two-fold increased risk of early-onset PrCa in a small clinical series and was unassociated with morbidity following treatment³⁶. In addition to the *ATM* region, we identified missense variants in three separate loci: rs2066827 within the cyclin-dependent kinase inhibitor *CDKN1B*, which controls cell cycle progression; rs33984059 within the transcription factor *RFX7*; and rs2277283 within *INCENP*, which encodes a centromere-interacting protein.

rs1048169 at 9p22 is located in the 3'UTR of *HAUS6* (**Figure 2, Panel B**), a gene that encodes a subunit of augmin, a protein complex required for proper microtubule formation and chromosome segregation during cell division³⁷. rs1048169 is also an eQTL for *HAUS6* expression. Interestingly, an additional lead SNP identified in this study, rs11666569 at 19p13, was an eQTL for two genes including *HAUS8*, which is another member of the augmin complex. These discoveries may implicate a potential role for augmin in PrCa susceptibility.

Variant rs7968403 (OR=1.06; $P=3.38 \times 10^{-12}$; **Figure 2, Panel C**) is situated within the first intron of *RASSF3*. Members of the Ras association domain family (RASSF) are putative tumor suppressors implicated in a range of biological processes³⁸. *RASSF3* is ubiquitously expressed across tissue types and has been observed to arrest the cell cycle in the G1 phase and induce apoptosis through the p53 pathway³⁹. A previously identified PrCa risk locus, ~100kb away, within the *RASSF6* family member was previously identified¹¹. However, rs7968403 was also an eQTL for the distant *WIF1* (WNT inhibitory factor 1) gene (**Figure 2, Panel C**). *WIF1* inhibits Wnt signaling and is frequently down regulated in PrCa⁴⁰, whilst aberrant activation of Wnt signaling is common in many solid tumor types. Restoration of *WIF1* expression has also been demonstrated to decrease cell motility and invasiveness in a metastatic PrCa cell-line and reduce tumor growth in a mouse xenograft model⁴¹. Both *RASSF3* and *WIF1* therefore represent plausible mechanisms for the modulation of PrCa risk at this locus.

rs28441558 at 17p13 is the lead variant for a cluster of highly correlated SNPs centered on the *CHD3* gene (**Figure 2, Panel D**). *CHD3* is an ATPase that forms a component of the NuRD (nucleosome remodeling and deacetylase) histone deacetylase complex, involved in chromatin remodeling. NuRD plays an important role in regulating gene expression, both as a silencer and activator of transcription, in addition to maintenance of genomic integrity and the DNA damage response⁴². Alterations to NuRD function have been implicated in several cancer types in a highly complex manner^{43,44}. Additionally however, rs28441558 was observed to be an eQTL for three genes; *LOC284023*, a currently uncharacterized non-coding RNA transcript, *GUCY2D*, a guanylate cyclase enzyme expressed predominantly in the retina and *ALOX15B*, a member of the lipoxygenase family of enzymes that produce fatty acid hydroperoxides.

Although *CHD3* appears to represent the most biologically plausible candidate gene for this locus, we cannot exclude a role for any of these genes.

Our pathway analysis based on mapping each SNP to the nearest gene (**see Online Methods**) using the meta-analysis summary association statistic identified several pathways implicated in PrCa susceptibility. The top 53 pathways detected (enrichment score, $ES > 0.50$) are provided (**Supplementary Table 15**). The most significant pathway detected was PD-1 signaling (ID: 389948), $ES = 0.74$, as defined by the REACTOME database (**Supplementary Figure 5**). This pathway is intriguing given the therapeutic potential of several checkpoint inhibitors focusing on the PD-1 signaling pathway to enhance immune responses⁴⁵.

In summary, we have identified 63 novel PrCa susceptibility variants, including strong candidate loci highlighting the DNA repair and cell cycle pathways. Previous studies likely overestimated the effect estimates of PrCa loci due to the “winner’s curse”, thus yielding a biased FRR and polygenic risk score (PRS). Here, we apply a weighted Bayesian correction approach and demonstrate our large sample size minimizes the “winner’s curse” bias (**Online Methods; Supplementary Figure 6**)⁴⁶. We applied the beta estimates calculated in our overall meta-analysis to the OncoArray sample set to calculate the FRR and PRS risk models (**Supplementary Table 16**). Our prediction models included 85 previously reported PrCa loci replicating in our overall meta-analysis and our 62 novel loci associated with overall PrCa risk. Assuming a familial risk estimate of 2.5 for PrCa^{47,48}, we demonstrate our 147 loci captures 28.4% of the FRR (**Supplementary Table 17**). The newly 62 identified PrCa loci increase the FRR by 4.4%. On the assumption of a log-additive model, the estimated RR for PrCa relative to men in the 25-75% PRS percentile (baseline group) was 5.71 (95%CI: 5.04-6.48) for men in the top 1% of the polygenic risk score (PRS) distribution and 2.69 (95%CI: 2.55-2.82) for individuals in the 90-99%-ile of the PRS (**Table 2**). The PRS score was positively associated with overall PrCa compared to all controls (OR=1.86; 95%CI: 1.83-1.89; **Supplementary Table 18**). Our novel associations highlight several biological pathways that suggest further investigation is warranted. The increased PRS can be used to improve the identification of men at high risk of PrCa and therefore inform PSA guidelines for screening and management to reduce the burden of over testing.

ACKNOWLEDGMENTS:

The authors wish to pay tribute to Brian Henderson, who was a driving force behind the OncoArray project, for his vision and leadership, and sadly passed away before seeing its fruition.

We thank the individuals who participated in these studies enabling this work.

Genotyping of the OncoArray was funded by the US National Institutes of Health (NIH) [U19 CA 148537 for ELucidating Loci Involved in Prostate cancer SuscEptibility (ELLIPSE) project and X01HG007492 to the Center for Inherited Disease Research (CIDR) under contract number HHSN268201200008I]. Additional analytic support was provided by NIH NCI U01 CA188392 (F.R.S.).

Funding for the iCOGS infrastructure came from: the European Community's Seventh Framework Programme under grant agreement n° 223175 (HEALTH-F2-2009-223175) (COGS), Cancer Research UK (C1287/A10118, C1287/A 10710, C12292/A11174, C1281/A12014, C5047/A8384, C5047/A15007, C5047/A10692, C8197/A16565), the National Institutes of Health (CA128978) and Post-Cancer GWAS initiative (1U19 CA148537, 1U19 CA148065 and 1U19 CA148112 - the GAME-ON initiative), the Department of Defence (W81XWH-10-1-0341), the Canadian Institutes of Health Research (CIHR) for the CIHR Team in Familial Risks of Breast Cancer, Komen Foundation for the Cure, the Breast Cancer Research Foundation, and the Ovarian Cancer Research Fund.

This work was supported by the Canadian Institutes of Health Research, European Commission's Seventh Framework Programme grant agreement n° 223175 (HEALTH-F2-2009-223175), Cancer Research UK Grants C5047/A7357, C1287/A10118, C1287/A16563, C5047/A3354, C5047/A10692, C16913/A6135, and The National Institute of Health (NIH) Cancer Post-Cancer GWAS initiative grant: No. 1 U19 CA 148537-01 (the GAME-ON initiative).

We would also like to thank the following for funding support: The Institute of Cancer Research and The Everyman Campaign, The Prostate Cancer Research Foundation, Prostate Research Campaign UK (now Prostate Action), The Orchid Cancer Appeal, The National Cancer Research Network UK, The National Cancer Research Institute (NCRI) UK. We are grateful for support of NIHR funding to the NIHR Biomedical Research Centre at The Institute of Cancer Research and The Royal Marsden NHS Foundation Trust.

The Prostate Cancer Program of Cancer Council Victoria also acknowledge grant support from The National Health and Medical Research Council, Australia (126402, 209057, 251533, , 396414, 450104, 504700, 504702, 504715, 623204, 940394, 614296,), VicHealth, Cancer Council Victoria, The Prostate Cancer Foundation of Australia, The Whitten Foundation, PricewaterhouseCoopers, and Tattersall's. EAO, DMK, and EMK acknowledge the Intramural Program of the National Human Genome Research Institute for their support.

The BPC3 was supported by the U.S. National Institutes of Health, National Cancer Institute (cooperative agreements U01-CA98233 to D.J.H., U01-CA98710 to S.M.G., U01-CA98216 to E.R., and U01-CA98758 to B.E.H., and Intramural Research Program of NIH/National Cancer Institute, Division of Cancer Epidemiology and Genetics).

CAPS GWAS study was supported by the Swedish Cancer Foundation (grant no 09-0677, 11-484, 12-823), the Cancer Risk Prediction Center (CRisP; www.crispcenter.org), a Linneus Centre (Contract ID 70867902) financed by the Swedish Research Council, Swedish Research Council (grant no K2010-70X-20430-04-3, 2014-2269)

PEGASUS was supported by the Intramural Research Program, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health.

For a full description of funding and acknowledgments see the **Supplementary Note**.

AUTHOR CONTRIBUTIONS:

Writing group: F.R.S., C.A.H., D.V.C., A. Al Olama, S.I.B., M.A, Z.K.J., R.A.E.
Conception and coordination of OncoArray synthesis: F.R.S., C.I.A., D.F.E., S.J.C., C.A.H., B.E.H., F.W. Database management: S.B., M.N.B., X.S., K.M., A.L.
Bioinformatics support: E.J.S., T.D., D.L., E.A., C.C.B., C.G. Genotyping calling and quality control: L.F., J.D., J.T. Provision of DNA samples and/or phenotypic data: V.L.S., S.M.G, B.D.C., C.M.T., P.J.G., I.M.T., J.B., S.C., L.M., J.C., L.H., W.T., G.P.R., H.G., M.A., T.N., P.P., N.P., J.S., T.L.J.T., C.S., A.A., D.A., S.W., A.W., N.H., C.M.L.W., A.M.D., N.B., L.A.M., E.G., G.L.A., O.C., G.C.T., S.K., L.E.B.F., K.D.S., T.F.O., M.B., L.M., E.M.G., D.E.N, J.L.D., F.C.H., R.M.M, R.C.T, T.J.K., R.J.H., N.E.F., A.F., S.A.I., M.C.S., B.S.R., S.L.K., H.O., Y.L., H.Z., N.F., X.M., X.G., G.W., Z.S., G.G.G., M.C.S., R.J.M., L.M.F., A.S.K., B.M.K., J.L., G.C.V., K.L.P., M.S., J.Y.P., T.A.S., H.L., J.L.S., C.C., D.W., J.L., E.A.O., M.S.G., B.G.N., S.F.N., M.W., R.B., M.A.R., P.I., H.B., K.C., B.H., C.M., M.L., T.S., J.K., C.J.L., E.M.J., M.R.T., P.P., M.C., S.L.N., L.S., Y.C.D., K.D.R., G.D.M., P.O., A.R., J.L., S.T., D.W.L., L.F.N., D.L., M.G., T.K., R.K., N.U., S.S., C.S., V.M., M.P., F.C., S.J., T.V.B., S.L., P.A.T., C.A.H., M.G.D., J.E.C., M.E.M., M.J.R., G.J., R.H.N.S, F.M., T.T., Y.A.K., J.X., K.K., L.C.A., H.P., A.M., S.N.T., S.K.M., D.J.S., S.L., C.T., J.M., D.J.H., E.R., A.S., F.C., L.N.K., L.L.M., R.N.H., M.J.M., Z.C., P.K., F.W., S.J.C., B.E.H., C.A.H., R.A.E. APCB investigators, IMPACT Study, Canary PASS investigators, BPC3, PRACTICAL, CAPS, PEGASUS, GAME-ON/ELLIPSE. All authors read and approved the final version of the manuscript.

COMPETING FINANCIAL INTERESTS STATEMENT:

The authors declare no competing financial interests.

REFERENCES:

1. Goh, C.L. *et al.* Genetic variants associated with predisposition to prostate cancer and potential clinical implications. *J Intern Med* **271**, 353-65 (2012).
2. Siegel, R.L., Miller, K.D. & Jemal, A. Cancer statistics, 2016. *CA Cancer J Clin* **66**, 7-30 (2016).
3. Cuzick, J. *et al.* Prevention and early detection of prostate cancer. *Lancet Oncol* **15**, e484-92 (2014).
4. Altekruse, S.F. *et al.* Spatial patterns of localized-stage prostate cancer incidence among white and black men in the southeastern United States, 1999-2001. *Cancer Epidemiol Biomarkers Prev* **19**, 1460-7 (2010).
5. Stanford, J.L. & Ostrander, E.A. Familial prostate cancer. *Epidemiol Rev* **23**, 19-23 (2001).
6. Bunker, C.H. *et al.* High prevalence of screening-detected prostate cancer among Afro-Caribbeans: the Tobago Prostate Cancer Survey. *Cancer Epidemiol Biomarkers Prev* **11**, 726-9 (2002).
7. Ghadirian, P., Howe, G.R., Hislop, T.G. & Maisonneuve, P. Family history of prostate cancer: a multi-center case-control study in Canada. *Int J Cancer* **70**, 679-81 (1997).
8. Gronberg, H., Damber, L. & Damber, J.E. Familial prostate cancer in Sweden. A nationwide register cohort study. *Cancer* **77**, 138-43 (1996).
9. Matikaine, M.P. *et al.* Relatives of prostate cancer patients have an increased risk of prostate and stomach cancers: a population-based, cancer registry study in Finland. *Cancer Causes Control* **12**, 223-30 (2001).
10. Eeles, R. *et al.* The genetic epidemiology of prostate cancer and its clinical implications. *Nat Rev Urol* **11**, 18-31 (2014).
11. Eeles, R.A. *et al.* Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. *Nat Genet* **45**, 385-91, 391e1-2 (2013).
12. Al Olama, A.A. *et al.* A meta-analysis of 87,040 individuals identifies 23 new susceptibility loci for prostate cancer. *Nat Genet* **46**, 1103-9 (2014).
13. Al Olama, A.A. *et al.* Multiple loci on 8q24 associated with prostate cancer susceptibility. *Nat Genet* **41**, 1058-60 (2009).
14. Amundadottir, L.T. *et al.* A common variant associated with prostate cancer in European and African populations. *Nat Genet* **38**, 652-8 (2006).
15. Eeles, R.A. *et al.* Identification of seven new prostate cancer susceptibility loci through a genome-wide association study. *Nat Genet* **41**, 1116-21 (2009).
16. Eeles, R.A. *et al.* Multiple newly identified loci associated with prostate cancer susceptibility. *Nat Genet* **40**, 316-21 (2008).
17. Gudmundsson, J. *et al.* Genome-wide association and replication studies identify four variants associated with prostate cancer susceptibility. *Nat Genet* **41**, 1122-6 (2009).
18. Gudmundsson, J. *et al.* Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nat Genet* **39**, 631-7 (2007).
19. Gudmundsson, J. *et al.* Common sequence variants on 2p15 and Xp11.22 confer susceptibility to prostate cancer. *Nat Genet* **40**, 281-3 (2008).

20. Gudmundsson, J. *et al.* Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. *Nat Genet* **39**, 977-83 (2007).
21. Haiman, C.A. *et al.* Genome-wide association study of prostate cancer in men of African ancestry identifies a susceptibility locus at 17q21. *Nat Genet* **43**, 570-3 (2011).
22. Kote-Jarai, Z. *et al.* Seven prostate cancer susceptibility loci identified by a multi-stage genome-wide association study. *Nat Genet* **43**, 785-91.
23. Schumacher, F.R. *et al.* Genome-wide association study identifies new prostate cancer susceptibility loci. *Hum Mol Genet* **20**, 3867-75.
24. Sun, J. *et al.* Evidence for two independent prostate cancer risk-associated loci in the HNF1B gene at 17q12. *Nat Genet* **40**, 1153-5 (2008).
25. Takata, R. *et al.* Genome-wide association study identifies five new susceptibility loci for prostate cancer in the Japanese population. *Nat Genet* **42**, 751-4 (2010).
26. Thomas, G. *et al.* Multiple loci identified in a genome-wide association study of prostate cancer. *Nat Genet* **40**, 310-5 (2008).
27. Yeager, M. *et al.* Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat Genet* **39**, 645-9 (2007).
28. Duggan, D. *et al.* Two genome-wide association studies of aggressive prostate cancer implicate putative prostate tumor suppressor gene DAB2IP. *J Natl Cancer Inst* **99**, 1836-44 (2007).
29. Amin Al Olama, A. *et al.* A meta-analysis of genome-wide association studies to identify prostate cancer susceptibility loci associated with aggressive and non-aggressive disease. *Hum Mol Genet* **22**, 408-15 (2013).
30. Amos, C.I. *et al.* The OncoArray Consortium: A Network for Understanding the Genetic Architecture of Common Cancers. *Cancer Epidemiol Biomarkers Prev* **26**, 126-135 (2017).
31. Delaneau, O., Marchini, J. & Zagury, J.F. A linear complexity phasing method for thousands of genomes. *Nat Methods* **9**, 179-81 (2012).
32. Howie, B.N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* **5**, e1000529 (2009).
33. de Bakker, P.I. *et al.* Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum Mol Genet* **17**, R122-8 (2008).
34. Leongamornlert, D. *et al.* Frequent germline deleterious mutations in DNA repair genes in familial prostate cancer cases are associated with advanced disease. *Br J Cancer* **110**, 1663-72 (2014).
35. Mateo, J. *et al.* DNA-Repair Defects and Olaparib in Metastatic Prostate Cancer. *N Engl J Med* **373**, 1697-708 (2015).
36. Meyer, A. *et al.* ATM missense variant P1054R predisposes to prostate cancer. *Radiother Oncol* **83**, 283-8 (2007).
37. Sanchez-Huertas, C. & Luders, J. The augmin connection in the geometry of microtubule networks. *Curr Biol* **25**, R294-9 (2015).

38. Volodko, N., Gordon, M., Salla, M., Ghazaleh, H.A. & Baksh, S. RASSF tumor suppressor gene family: biological functions and regulation. *FEBS Lett* **588**, 2671-84 (2014).
39. Kudo, T. *et al.* The RASSF3 candidate tumor suppressor induces apoptosis and G1-S cell-cycle arrest via p53. *Cancer Res* **72**, 2901-11 (2012).
40. Wissmann, C. *et al.* WIF1, a component of the Wnt pathway, is down-regulated in prostate, breast, lung, and bladder cancer. *J Pathol* **201**, 204-12 (2003).
41. Yee, D.S. *et al.* The Wnt inhibitory factor 1 restoration in prostate cancer cells was associated with reduced tumor growth, decreased capacity of cell migration and invasion and a reversal of epithelial to mesenchymal transition. *Mol Cancer* **9**, 162 (2010).
42. Allen, H.F., Wade, P.A. & Kutateladze, T.G. The NuRD architecture. *Cell Mol Life Sci* **70**, 3513-24 (2013).
43. Lai, A.Y. & Wade, P.A. Cancer biology and NuRD: a multifaceted chromatin remodelling complex. *Nat Rev Cancer* **11**, 588-96 (2011).
44. Basta, J. & Rauchman, M. The nucleosome remodeling and deacetylase complex in development and disease. *Transl Res* **165**, 36-47 (2015).
45. McDermott, D.F. & Atkins, M.B. PD-1 as a potential target in cancer therapy. *Cancer Med* **2**, 662-73 (2013).
46. Zhong, H. & Prentice, R.L. Bias-reduced estimators and confidence intervals for odds ratios in genome-wide association studies. *Biostatistics* **9**, 621-34 (2008).
47. Kicinski, M., Vangronsveld, J. & Nawrot, T.S. An epidemiological reappraisal of the familial aggregation of prostate cancer: a meta-analysis. *PLoS One* **6**, e27130 (2011).
48. Albright, F. *et al.* Prostate cancer risk prediction based on complete prostate cancer family history. *Prostate* **75**, 390-8 (2015).

FIGURE LEGENDS:

Figure 1. ELLIPSE/PRACTICAL study overview of prostate cancer (PrCa) GWAS meta-analysis. The top section describes the PrCa GWAS meta-analysis published in 2014 (AA Olama *et al*, Nature Genetics 2014¹²) where 23 novel variants were identified. The current PrCa GWAS meta-analysis incorporates an additional 46,939 PrCa cases and 27,910 controls independent of the meta-analyses. The current meta-analysis discovered 62 novel variants associated with overall PrCa and 1 novel variant associated with early-onset PrCa.

Figure 2. Locus Explorer plots depicting the statistical association with PrCa and biological context of variants from four of the novel prostate cancer loci identified (N=74,849 biologically independent samples). For each panel (a-d), top panels depict Manhattan plots of variant $-\log_{10} P$ values (y-axis), with the index SNP labeled. Variants that were directly genotyped by the OncoArray are represented as triangles and imputed variants are represented as circles. Variants in linkage disequilibrium with the index SNP are denoted by color (red = $r^2 > 0.8$, orange = $r^2 0.6-0.8$, yellow = $r^2 0.4-0.6$, green = $r^2 0.2-0.4$, blue = $r^2 \leq 0.2$). Middle panels depict the relative locations of selected biological annotations; histone marks within 7 cell lines from the ENCODE project; genes for which the index SNP is an eQTL in the TCGA prostate adenocarcinoma dataset; chromatin state annotation by ChromHMM in PrEC cells; conserved elements within the genome and DNaseI hypersensitivity sites in ENCODE prostate cell lines. The lower panel denotes the position of genes within the region, with genes on the positive and negative strands marked in green and purple, respectively. The horizontal axis represents genomic co-ordinates in the hg19 reference genome. **(a)** rs1800057 (chr11:107643000-108644000) - The index variant is a non-synonymous SNP in the *ATM* gene. **(b)** rs1048160 (chr9:18556000-19557000) - The index variant is located within the 3'UTR of the *HAUS6* gene and is an eQTL for *HAUS6*. **(c)** rs7968403 (chr12:64513000-65514000) - The signal is centered on the *RASSF3* gene, with the index variant located within the first intron. This SNP is also situated within a region annotated for multiple regulatory markers and is an eQTL for the more distant *WIF1* gene. **(d)** rs28441558 (chr17:7303000-8304000) - The signal implicates a cluster of highly correlated variants centered upon the *CHD3* gene. The index SNP is also an eQTL for three other more distantly located genes.

Table 1. Prostate cancer OncoArray and GWAS meta-analysis for 63 novel regions.

SNP	Reference RAF ^a	Band	Position	Nearest Gene	Alleles ^b	RAF ^c	OR ^d	95% CI ^e	P-value ^f
Novel loci associated with overall prostate cancer									
rs56391074	0.329	1p22.3	88210715	RP11-60A14.1	AT/A	0.38	1.05	1.03-1.06	1.7E-08
rs34579442	0.316	1q21.3	153899900	DENND4B	C/CT	0.34	1.07	1.05-1.09	4.5E-14
rs62106670	0.400	2p25.1	8597123	AC011747.3	T/C	0.38	1.05	1.04-1.07	7.1E-09
rs74702681	0.024	2p14	66652885	MEIS1-AS3	T/C	0.02	1.17	1.11-1.23	2.0E-09
rs11691517	0.750	2q13	111893096	BCL2L11	T/G	0.74	1.07	1.05-1.08	3.5E-12
rs34925593	0.481	2q31.1	174234547	CDCA7	C/T	0.48	1.05	1.03-1.07	2.8E-08
rs59308963	0.726	2q33.1	202123479	CASP8	T/TATTCTGTC	0.73	1.05	1.03-1.07	2.4E-08
rs1283104	0.407	3q13.12	106962521	DUBR	G/C	0.38	1.05	1.03-1.07	8.8E-09
rs182314334	0.888	3q25.1	152004202	MBNL1	T/C	0.90	1.09	1.06-1.12	4.1E-11
rs142436749	0.012	3q26.2	169093100	MECOM	G/A	0.01	1.25	1.16-1.34	4.7E-09
rs10793821	0.580	5q31.1	133836209	RNU6-456P	T/C	0.57	1.05	1.04-1.07	5.4E-11
rs76551843	0.991	5q35.1	169172133	DOCK2	A/G	0.99	1.31	1.19-1.44	1.7E-08
rs4976790	0.096	5q35.3	177968915	COL23A1	T/G	0.11	1.08	1.05-1.10	6.7E-09
rs12665339	0.148	6p21.33	30601232	ATAT1	G/A	0.17	1.06	1.04-1.08	5.6E-09
rs9296068	0.645	6p21.32	32988695	HLA-DOA	T/G	0.65	1.05	1.03-1.07	1.3E-08
rs9469899	0.356	6p21.31	34793124	UHRF1BP1	A/G	0.36	1.05	1.03-1.07	5.3E-09
rs4711748	0.232	6p21.1	43694598	RP1-261G23.5	T/C	0.23	1.05	1.03-1.07	3.4E-08
rs527510716	0.251	7p22.3	1944537	MAD1L1	C/G	0.24	1.06	1.04-1.08	4.9E-08
rs11452686	0.567	7p21.1	20414110	ITGB8	T/TA	0.56	1.05	1.03-1.07	7.8E-09
rs17621345	0.758	7p14.1	40875192	SUGCT	A/C	0.74	1.07	1.05-1.09	6.7E-14
rs1048169	0.367	9p22.1	19055965	HAUS6	C/T	0.38	1.06	1.05-1.08	6.5E-14
rs10122495	0.296	9p13.3	34049779	RN7SKP114	T/A	0.31	1.05	1.03-1.07	1.3E-08

rs1182	0.258	9q34.11	132576060	TOR1A	A/C	0.22	1.06	1.04-1.08	1.1E-09
rs141536087	0.166	10p15.3	854691	LARP4B	GCGCA/G	0.15	1.08	1.06-1.11	9.0E-13
rs1935581	0.605	10q23.31	90195149	RNLS	C/T	0.63	1.05	1.03-1.07	6.5E-09
rs7094871	0.540	10q25.2	114712154	TCF7L2	G/C	0.54	1.04	1.03-1.06	4.8E-08
rs1881502	0.193	11p15.5	1507512	MOB2	T/C	0.19	1.06	1.04-1.08	7.4E-09
rs61890184 ^g	0.088	11p15.4	7547587	PPFIBP2	A/G	0.12	1.07	1.05-1.10	6.6E-09
rs547171081	0.468	11p11.2	47421962	RP11-750H9.5	CGG/C	0.47	1.05	1.03-1.07	3.4E-08
rs2277283	0.300	11q12.3	61908440	INCENP	C/T	0.31	1.06	1.04-1.08	3.0E-10
rs12785905	0.051	11q13.2	66951965	KDM2A	C/G	0.05	1.12	1.08-1.17	7.8E-09
rs11290954	0.688	11q13.5	76260543	C11orf30	AC/A	0.68	1.06	1.05-1.08	7.4E-13
rs1800057	0.031	11q22.3	108143456	ATM	G/C	0.02	1.16	1.10-1.22	8.1E-09
rs138466039	0.009	11q24.2	125054793	PKNOX2	T/C	0.01	1.32	1.22-1.44	2.0E-11
rs878987	0.143	11q25	134266372	B3GAT1	G/A	0.15	1.07	1.04-1.09	4.8E-08
rs2066827	0.757	12p13.1	12871099	CDKN1B	T/G	0.76	1.06	1.04-1.08	2.3E-09
rs10845938	0.554	12p13.1	14416918	RNU6-491P	G/A	0.55	1.06	1.04-1.08	9.8E-13
rs7968403	0.655	12q14.2	65012824	RASSF3	T/C	0.64	1.06	1.04-1.08	3.4E-12
rs5799921	0.697	12q21.33	90160530	RNU6-148P	GA/G	0.68	1.06	1.04-1.08	7.0E-12
rs7295014	0.342	12q24.33	133067989	FBRSL1	G/A	0.35	1.05	1.04-1.07	9.5E-10
rs1004030	0.581	14q11.2	23305649	MMP14	T/C	0.58	1.05	1.03-1.06	1.5E-08
rs11629412	0.582	14q13.3	37138294	PAX9	C/G	0.58	1.06	1.04-1.08	2.3E-12
rs4924487	0.836	15q15.1	40922915	CASC5	C/G	0.81	1.06	1.04-1.09	1.3E-08
rs33984059	0.982	15q21.3	56385868	RFX7	A/G	0.98	1.19	1.12-1.27	1.1E-08
rs112293876	0.280	15q22.31	66764641	MAP2K1	C/CA	0.29	1.06	1.04-1.08	3.5E-10
rs11863709	0.945	16q21	57654576	GPR56	C/T	0.96	1.16	1.11-1.21	1.8E-11
rs201158093	0.435	16q23.3	82178893	RP11-510J16.5	TAA/TA	0.44	1.05	1.03-1.07	9.1E-09
rs28441558	0.050	17p13.1	7803118	CHD3	C/T	0.05	1.16	1.12-1.20	1.0E-16
rs142444269	0.798	17q11.2	30098749	RP11-805L22.3	C/T	0.78	1.07	1.05-1.09	3.2E-10

rs2680708	0.623	17q22	56456120	RNF43	G/A	0.61	1.05	1.03-1.06	1.6E-08
rs8093601	0.459	18q21.2	51772473	MBD2	C/G	0.44	1.05	1.03-1.06	2.3E-08
rs28607662	0.085	18q21.2	53230859	TCF4	C/T	0.10	1.08	1.05-1.11	2.8E-08
rs12956892	0.300	18q21.32	56746315	OACYLP	T/G	0.30	1.05	1.03-1.07	7.7E-09
rs533722308	0.390	18q21.33	60961193	BCL2	CT/C	0.42	1.05	1.03-1.07	1.2E-08
rs10460109	0.414	18q22.3	73036165	TSHZ1	T/C	0.42	1.05	1.03-1.06	3.5E-08
rs11666569	0.728	19p13.11	17214073	MYO9B	C/T	0.71	1.05	1.03-1.07	8.2E-09
rs118005503	0.912	19q12	32167803	THEG5	G/C	0.91	1.09	1.06-1.13	7.3E-09
rs61088131	0.848	19q13.2	42700947	POU2F2	T/C	0.82	1.06	1.04-1.09	8.8E-09
rs11480453	0.641	20q11.21	31347512	DNMT3B	C/CA	0.60	1.05	1.03-1.06	3.2E-08
rs6091758	0.465	20q13.2	52455205	BCAS1	G/A	0.47	1.07	1.06-1.09	6.4E-18
rs9625483	0.026	22q12.1	28888939	TTC28	A/G	0.03	1.14	1.09-1.20	2.4E-08
rs17321482	0.873	23p22.2	11482634	ARHGAP6	C/T	0.87	1.07	1.05-1.09	2.1E-13
Novel loci associated with early-onset									
rs138004030	0.920	6q27	170475879	LOC154449	G/A	0.91	1.27	1.17-1.38	2.9E-08

^a Risk allele frequency - 1000 Genomes Project Europeans

^b Risk allele/Reference allele

^c Risk allele frequency

^d Odds ratio and confidence interval

^e Confidence interval

^f P-values are generated from a likelihood ratio test

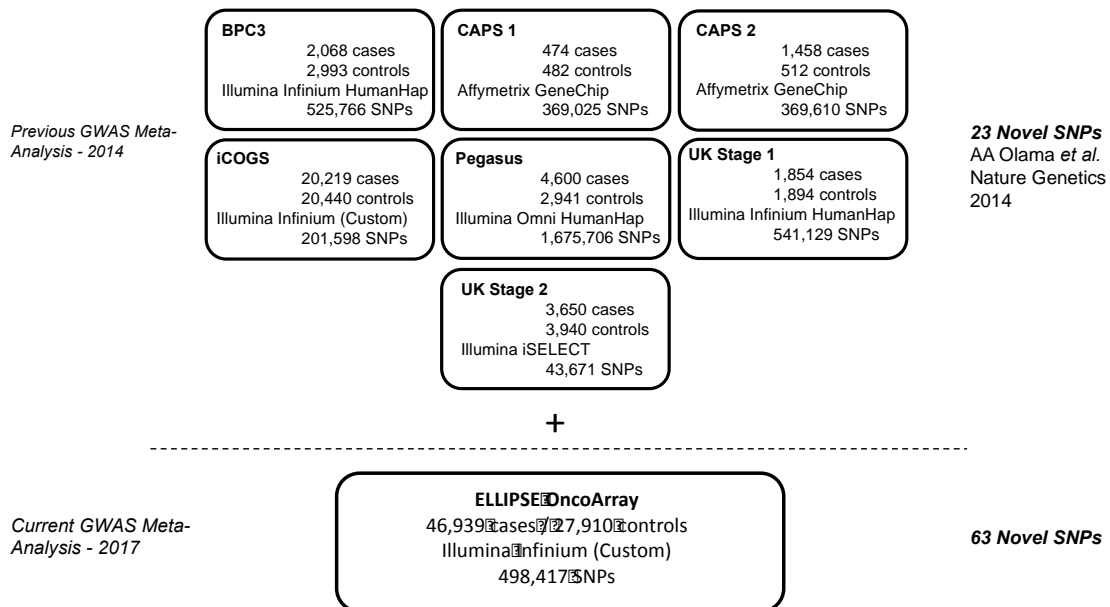
^g Region previously reported by Wang et al (Nat Comm 2015), rs12791447; rs61890184-rs12791447 r2 (EUR)=0.41

Table 2. Polygenic Risk Score (PRS) estimation using 147 prostate cancer susceptibility variants.

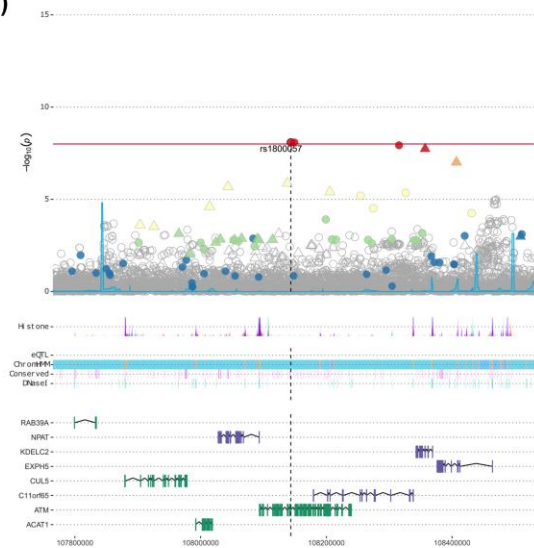
Risk Category Percentiles ^a	RR	95% CI
<1%	0.15	0.11-0.20
1-10%	0.35	0.32-0.37
10-25%	0.54	0.51-0.57
25-75%	1.00 (Baseline)	
75-90%	1.74	1.67-1.82
90-99%	2.69	2.55-2.82
≥99%	5.71	5.04-6.48

^a Polygenetic Risk Score (PRS) percentiles based on the cumulative score distributed among controls. The beta coefficients computed from the European overall meta-analysis was applied to determine the PRS risk among individuals in the OncoArray study.

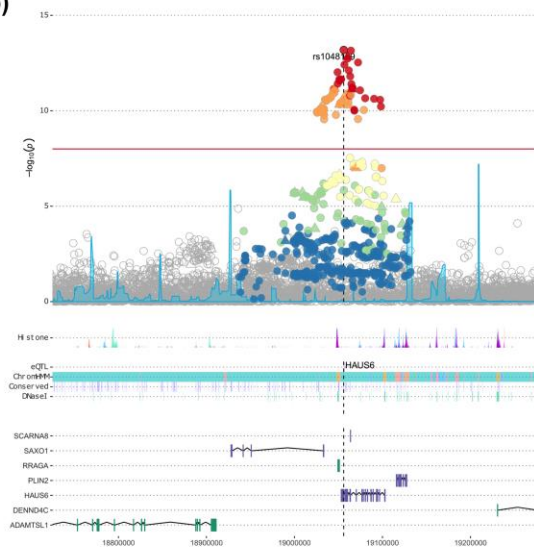
^b Relative risk and 95% confidence intervals.



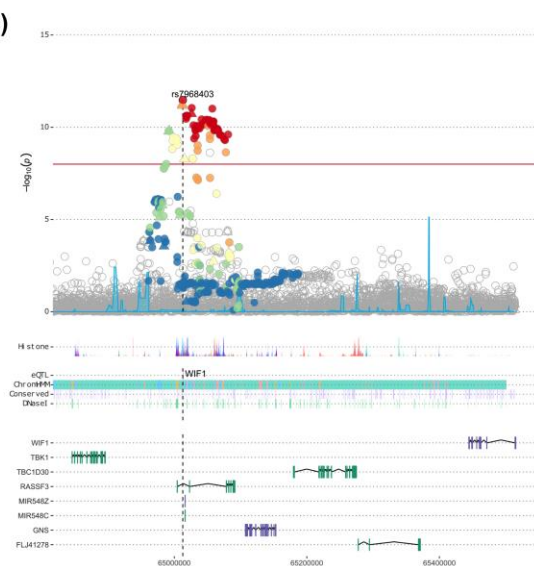
(a)



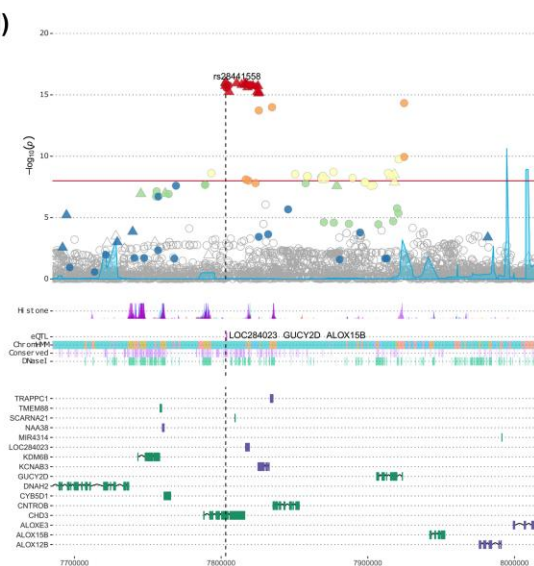
(b)



(c)



(d)



ONLINE METHODS:

Study subjects

A brief overview and study details for participating prostate cancer (PrCa) studies in the newly genotyped OncoArray project are provided in **Supplementary Table 1**¹ for men of European ancestry. All studies were approved by the appropriate ethics committees and informed consent was obtained from all participants. **Supplementary Table 2** summarizes the PrCa sample series of the Elucidating Loci Involved in Prostate Cancer Susceptibility (ELLIPSE) consortia contributing both newly obtained genotyping data for the OncoArray and previous genome-wide association studies (GWAS). The majority of the studies contributing to the OncoArray were case-control studies primarily based in either the United States or Europe. In total 52 new studies provided core data on disease status, age at diagnosis (age at observation or questionnaire for controls), family history of PrCa, and clinical factors for cases (e.g. PSA at diagnosis, Gleason score, etc.) for 48,455 PrCa cases and 28,321 disease-free controls. Previous GWAS contributed an additional 32,255 PrCa cases and 33,202 disease-free controls of European ancestry for the overall meta-analysis¹. **Supplementary Table 3** provides quality control information by consortia (*i.e.* OncoArray project, UK GWAS, etc) for both samples and SNPs. After removing all overlapping samples the OncoArray contribution for newly genotyped samples was 46,939 PrCa cases and 27,910 disease-free controls.

Several strata-specific analyses were implemented to evaluate the impact of genetic variation in PrCa disease aggressiveness. **Supplementary Table 4** describes the analysis title, outcome and reference groups, and the statistical model used. Several classification schemes (*i.e.* low aggressiveness, intermediate aggressiveness, etc.) were implemented to better assess the spectrum of genetic involvement. All classification schemes incorporated the diagnostic clinical features PSA, tumor stage and Gleason score. In order to compare to previous PrCa aggressive analyses¹ by our research group, we included the 'Advanced (plus death due to PrCa)' classification. Contributing study groups missing clinical features were excluded (**Supplementary Table 2**). Individuals with missing or granular clinical information were excluded. The strata-specific sample sizes by PrCa GWAS consortium are

provided in **Supplementary Table 5**. Furthermore we analysed Gleason score as a continuous variable.

OncoArray SNP selection

The NCI Genetic Associations and Mechanisms in Oncology (GAME-ON) consortia (<http://epi.grants.cancer.gov/gameon/>) provided SNPs to be included on the Illumina OncoArray. Approximately 50% of the OncoArray was a compilation of SNP lists by the GAME-ON disease consortia of cancer (breast, colorectal, lung, ovarian, and prostate), a common set of variants for common risk regions, other related traits (*i.e.* BMI, age at menarche, etc), pharmacogenetics, and candidates². The remaining content of the OncoArray was selected as a “GWAS backbone” (Illumina HumanCore), which aimed to provide high coverage for the majority of common variants through imputation. Approximately 79k SNPs were selected specifically for their relevance to PrCa, based on prior evidence of association with overall or subtype-specific disease, fine-mapping of known PrCa regions, and candidate submissions (*i.e.* survival, exome sequencing, etc). In order to maximize efficiency of the array, cancer-specific candidate lists were merged to remove redundant genetic variation³.

Genotype calling and quality control

Details of the genotype calling and quality control (QC) for the iCOGS and GWAS are described elsewhere⁴⁻²⁰.

Of the 568,712 variants selected for genotyping on OncoArray, 533,631 were successfully manufactured on the array (including 778 duplicate probes). OncoArray genotyping of ELLIPSE studies was conducted at five sites (Cambridge [UK], CIDR, Copenhagen, USC, NCI). Details of the genotyping calling for the OncoArray are described in more detail elsewhere³. Briefly, we developed a single calling pipeline that was applied to more than 500,000 samples across the GAME-ON consortia. An initial cluster file was generated using 56,284 samples selected from all major genotyping centers and ethnicities, using the Gentrain2 algorithm. Variants likely to have problematic clusters were selected for manual inspection using the following criteria: call rate below 99%, minor allele frequency (MAF) <0.001, poor Illumina intensity and clustering

metrics, deviation from the MAF observed in the 1000 Genomes Project (1KGP) using the criterion: $\frac{(|p_1 - p_0| - 0.01)^2}{(p_1 + p_0)(2 - p_1 - p_0)} > C$, where p_0 and p_1 are the minor frequencies in the 1KGP and OncoArray datasets, respectively, and $C=0.008$. This resulted in manual adjustment of the cluster file for 3,964 variants, and the exclusion of 16,526 variants. The final cluster file was then applied to the full dataset.

Our quality control pipeline for ELLIPSE excluded SNPs with a call rate <95% by study, not in Hardy-Weinberg equilibrium ($P < 10^{-7}$ in controls, or $P < 10^{-12}$ in cases) or with concordance <98% among 11,260 duplicate pairs. In order to minimize imputation errors, we additionally excluded SNPs with a MAF <1% and a call rate <98% in any study, SNPs that could not be linked to the 1KGP reference, those with MAF for Europeans that differed from that for the 1KGP and a further 16,526 SNPs where the cluster plot was judged to be not ideal. Of the 533,631 manufactured SNPs on the OncoArray, we retained 498,417 SNPs among our samples of European ancestry following QC.

We excluded duplicate samples and first-degree relatives within each study, duplicates across studies, samples with a call rate <95%, and samples with extreme heterozygosity (>4.9 standard deviations from the mean for the reported ethnicity). We excluded duplicated samples as well as first-degree relatives across the GWAS studies CAPS1, CAPS2, UK Stage 1, UK Stage 2, and iCOGS. Duplicate and first-degree related samples were assessed across the BPC3 and Pegasus GWAS studies as well. Ancestry was computed using a principal component analysis using 2,318 informative markers on a subset of ~47,000 samples and projected onto the complete OncoArray dataset. The current analysis was restricted to men of European ancestry, defined as individuals with an estimated proportion of European ancestry >0.8, with reference to the HapMap populations, based on the first two principal components. Of the 78,182 samples genotyped (regardless of race/ethnicity), the final dataset consisted of 74,849 samples, of which 46,939 PrCa cases and 27,910 disease-free controls (**Supplementary Table 3**) after excluding overlap samples, were meta-analysed with previous studies.

Imputation

Genotypes for ~70M SNPs were imputed for all samples using the October 2014 (Phase 3) release of the 1KGP data as the reference panel. The OncoArray and GWAS datasets were imputed using a two-stage imputation approach, using SHAPEIT²¹ for phasing and IMPUTEv2²² for imputation. The imputation was performed in 5Mb non-overlapping intervals. All subjects were split into subsets of ~10,000 samples, with subjects from the same group in the subset. We imputed genotypes for all SNPs that were polymorphic (MAF>0.1%) in European samples. We excluded data for all monomorphic SNPs and those with an imputation $r^2 < 0.3$ leaving a total of 20,370,935 SNP across chromosomes 1-22 and chromosome X. Of the SNPs imputed, 49.3% had a MAF<1%, 15.2% had a MAF ranging between 1-5%, and 35.5% had a MAF≥5%.

Statistical analyses

Per-allele odds ratios and standard errors were generated for the OncoArray and each GWAS, adjusting for principal components and study relevant covariates using logistic regression. The OncoArray and iCOGS analyses were additionally stratified by country and study, respectively. We used the first seven principal components in our analysis of individuals of European ancestry, as additional components did not further reduce inflation in the test statistics.

Odds ratio (OR) estimates were derived using either SNPTTEST (https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html) or an in-house C++ program (**Supplementary Table 3**). OR estimates and standard errors were combined by a fixed effects inverse variance meta-analysis using METAL²³. All statistical tests conducted were two-sided.

Our analyses included overall PrCa and several clinically relevant strata. These included: 1) high vs low aggressive PrCa; 2) high vs low/intermediate aggressive PrCa; 3) advanced vs non-advanced PrCa; 4) advanced PrCa vs controls; 5) early-onset PrCa (≤55 yrs) vs controls; and 6) Gleason score (**Supplementary**

Tables 4 & 5). We defined low aggressive as tumor stage $\leq T1$ and Gleason ≤ 6 and PSA <10 ng/mL, intermediate aggressive as tumor stage T2 or Gleason=7 or PSA 10-20 ng/mL, high aggressive as tumor stage T3/T4 or N1 or M1 or Gleason ≥ 8 or PSA >20 ng/mL, and advanced as either metastatic disease, Gleason ≥ 8 , PSA >100 or PrCa-related deaths (**Supplementary Table 4**).

Definition of new hits

To search for novel loci, we assessed all SNPs excluding those within a known PrCa locus, defined by current fine-mapping assessments (**Supplementary Table 7**). SNPs that were associated with disease risk at $P < 5 \times 10^{-8}$ in the meta-analysis (GWAS and OncoArray) were considered novel. The SNP with the lowest p-value in a region was considered the lead SNP. Imputation quality assessed by IMPUTE2 imputation r^2 in the OncoArray dataset (**Supplementary Table 8**).

For ten regions where the newly identified locus was near a previously known region, we reported a novel association if the pairwise r^2 between the new and the previously known SNP was less than 0.2. For novel PrCa associations where the variant was imputed in the OncoArray study samples series and had an imputed quality score less than 0.70, we assessed the quality of the imputation by masking the variant in a subset of the 1KGP European sample and calculating the concordance following re-imputation in the remaining 1KGP samples.

Reliability of Imputation

Novel SNPs with an IMPUTE2 $r^2 < 0.80$ among the OncoArray sample series (**Supplementary Table 8**) were flagged for further investigation to minimize the probability of a false positive. First, we examined linkage disequilibrium (LD) plots (<http://locuszoom.org/>) for poorly imputed SNPs (± 500 kb) including only genotyped SNPs within the region. The imputed index SNP was included in the plot to determine the strength of LD with nearby signals and assess a pattern of association. Furthermore, we performed an imputation experiment using the

2,504 1KGP Phase 3 samples. We split this sample into two parts: a random sample of 259 individuals of European ancestry (excluding the Finnish) and a mixed-population reference panel of 2,245 individuals. The random sample of 259 individuals of European ancestry was filtered to include only the genetic variants available from the OncoArray following QC. This ensured the same imputation input used in the overall imputation. The 259 individuals were imputed using 2,245 individuals as the reference panel. A 5 MB segment of the genome was selected based on the target SNP (+/- 250 MB). SHAPEIT2 was used for pre-phasing and IMPUTE2 for imputation. Customized imputation settings included an effect size of 20,000, allowance of large region imputation and a random seed of 12345. A weighted linear Kappa statistic was calculated to determine correlation of the imputation with the true genotypes.

We evaluated four SNPs where the IMPUTE2 r^2 was less than 0.80 in the OncoArray sample series: rs527510716 (Chr 7), rs6602880 (Chr 10), rs533722308 (Chr 18) and rs144166867 (Chr X). **Supplementary Figure 3** includes the LD plots for three of the poorly imputed SNPs. The variant rs144166867 (Chr X) could not be plotted given no genotype SNPs were available +/-500 KB on the OncoArray. Both LD plots for markers rs527510716 (Chr 7) and rs533722308 (Chr 18) showed significant associations ($P < 1 \times 10^{-3}$) for several genotype markers with moderate LD of the index SNP. The Kappa coefficient for markers rs527510716 (Chr 7) and rs533722308 (Chr 18) was 0.911 and 0.931, respectively (**Supplementary Table 9**). The marker rs6602880 (Chr 10) had a Kappa coefficient of 0.812 and was the only significant variant in the LD plot. The Kappa coefficient for marker rs144166867 (Chr X) was 0.665 (**Supplementary Table 9**). The markers rs6602880 (Chr 10) and rs144166867 (Chr X) are most likely false positives due to poor imputation for these regions.

Proportion of familial risk explained

The contribution of the known SNPs to the familial risk of PrCa, under a multiplicative model, was computed using the formula

$$\frac{\sum_k (\log I_k)}{\sum (\log I_0)}$$

where λ_0 is the observed familial risk to first degree relatives of PrCa cases^{24,25}, assumed to be 2.5, and λ_k is the familial relative risk due to locus k, given by:

$$\lambda_k = \frac{p_k r_k^2 + q_k}{(p_k r_k + q_k)^2}$$

where p_k is the frequency of the risk allele for locus k, $q_k = 1 - p_k$ and r_k is the estimated per-allele odds ratio.

Based on the assumption of a log-additive model, we constructed a polygenic risk score (PRS) from the summed risk allelic dosages weighted by the per-allele log-odds ratios. Thus for each individual j we derived:

$$Score_j = \sum_{i=1}^N \beta_i g_{ij}$$

Where:

N : Number of SNPs

g_{ij} : Allele dose at SNP_i for individual j

β_i : Per-allele log-odds ratio of SNP_i

The risk of PrCa was estimated for the percentile of the distribution of the PRS (<1%, 1-10%, 10-25%, 25-75%, 75-90%, 90-99%, >99% and <10%, 10-25%, 25-75%, 75-90%, >90%) where cumulative score thresholds were determined by the observed distribution among controls. We applied effect sizes and allele frequencies obtained from the overall meta-analysis of Europeans to estimate risk scores for individuals of European ancestry in the OncoArray study²⁶. A standardized PRS score was calculated by dividing the observed PRS score by the standard deviation of the PRS score among controls. A logistic regression framework was used to evaluate the percentile comparisons and determine the risk estimate. The models were adjusted for the first seven principal components to account for population stratification and stratified by country.

The FRR and PRS risk estimation was limited to the variants where our overall meta-analysis observed a statistically significant association. In total, we included

147 PrCa index SNPs in our risk score modelling, including 85 previously published associations and the 62 novel findings reported here. To correct for potential bias in effect estimation of newly discovered variants, we implemented a fully Bayesian version of a weighted correction given in Zhong and Prentice, Eq 3.4²⁷. Specifically, we place a normal prior distribution on MLE effect estimates of the form $\beta_m \sim N(\beta_{Cor}, \tau^2)$. Here, β_m is the log odds ratio from the overall meta-analysis; β_{Cor} is the bias corrected estimate calculated using the expectation-adjusted estimator from Eq 3.1 in Zhong and Prentice; and τ is a pre-specified variance of the effect distribution reflecting the bias and is defined as $\tau = |\hat{\beta}_m - \beta_{Cor}|$.

eQTL analyses

Genotype and gene expression data were downloaded from The Cancer Genome Atlas (TCGA) for 494 samples with PrCa (<https://gdc-portal.nci.nih.gov>). Quality Control (QC) was performed on both these datasets as follows: on the genotype, we filtered out samples with high heterozygosity (mean heterozygosity +/- 2 standard deviation) and missing genotypes, duplicated or related samples. We then performed Principal Component Analysis on the 494 samples plus 2,506 samples from 1KGP to infer the ancestry of the TCGA samples; samples of non-European ancestry were removed. We also filtered out variants with missing call rate > 5%. For the expression data, samples from two plates had, on average, much higher expression values than the remaining samples, and these were excluded. We also filtered genes with mean expression across samples <= 6 counts. Finally, expression values were quantile-normalized by samples and rank-transformed by genes. After QC we used the data from 359 samples. For the eQTL analysis, 35 PEER factors from the top 10,000 expressed genes were used as covariates, plus three genotyping PCs (which explained 18% of total variation). eQTL analysis was performed using FastQTL with 1,000 permutations over the 85 regions. We used a window of 1 Megabases (upstream/downstream) from the transcription start site (TSS) of each gene.

Gene Set Enrichment Analyses

The file Human_GOBP_AllPathways_no_GO_iea_September_01_2016_symbol.gmt (http://baderlab.org/EM_GeneSets), from the GeneSets database²⁸, was used for all analyses. This database contains pathways from Reactome²⁹, NCI Pathway Interaction Database³⁰, GO (Gene Ontology) biological process³¹, HumanCyc³², MSigdb³³, NetPath³⁴ and Panther³⁵. We manually corrected several pathways where the *PDPK1* gene was entered as *PDK1*. GO pathways inferred from electronic annotation terms were excluded. The same pathway (e.g. apoptosis) may be defined in two or more databases with potentially different sets of genes, and all versions of these duplicate/overlapping pathways were included. Pathway size was determined by the total number of genes in the pathway to which SNPs in the imputed GWAS dataset could be mapped. To provide more biologically meaningful results, and reduce false positives, only pathways that contained between 10 and 200 genes were considered.

Gene information (hg19) was downloaded from the ANNOVAR³⁶ website (<http://www.openbioinformatics.org/annovar/>). SNPs were mapped to the nearest gene within 500kb window; those that were further away from any gene were excluded. Gene significance was calculated by assigning the lowest p-value observed across all SNPs assigned to a gene^{37,38}, based on the combined European meta-analysis (previous GWAS and OncoArray).

The gene set enrichment analysis (GSEA)²⁸ algorithm, as implemented in the GenGen package (<http://gengen.openbioinformatics.org/en/latest/>)^{38,39} was used to perform pathway analysis. Briefly, the algorithm calculates an enrichment score (ES) for each pathway based on a weighted Kolmogorov-Smirnov statistic³⁹. To calculate the ES we performed 100 permutations and averaged the final score. Pathways that have most of their genes at the top of the ranked list of genes obtain higher ES values. Only pathways with positive ES and at least one gene with $P < 5 \times 10^{-8}$ were retained for subsequent analysis. An enrichment map

was created using the Enrichment Map (EM) v 2.1.0 app²⁸ in Cytoscape v3.40⁴⁰, applying force directed layout, weighted mode. We restricted our pathway analysis those with an $ES \geq 0.50$ to ensure a true positive rate > 0.20 and a false positive rate < 0.15 .

DATA AVAILABILITY:

The OncoArray genotype data and relevant covariate information (*i.e.* ethnicity, country, principal components, etc.) generated during this study will be deposited into dbGAP for access. In total 47 of the 52 OncoArray studies encompassing nearly 90% of the individual samples will be available (**Supplementary Table 19**). The previous meta-analysis summary results and genotype data currently¹² are available in dbGAP (Accession #: phs001081.v1.p1). The complete meta-analysis summary associations statistics is publicly available at the PRACTICAL website (<http://practical.icr.ac.uk/blog/>).

METHODS ONLY REFERENCES

49. Al Olama, A.A. *et al.* A meta-analysis of 87,040 individuals identifies 23 new susceptibility loci for prostate cancer. *Nat Genet* **46**, 1103-9 (2014).
50. Amos, C.I. *et al.* The OncoArray Consortium: A Network for Understanding the Genetic Architecture of Common Cancers. *Cancer Epidemiol Biomarkers Prev* **26**, 126-135 (2017).
51. Amos, C.I. *et al.* The OncoArray Consortium: a Network for Understanding the Genetic Architecture of Common Cancers. *Cancer Epidemiol Biomarkers Prev* (in press).
52. Al Olama, A.A. *et al.* Multiple loci on 8q24 associated with prostate cancer susceptibility. *Nat Genet* **41**, 1058-60 (2009).
53. Amundadottir, L.T. *et al.* A common variant associated with prostate cancer in European and African populations. *Nat Genet* **38**, 652-8 (2006).
54. Duggan, D. *et al.* Two genome-wide association studies of aggressive prostate cancer implicate putative prostate tumor suppressor gene DAB2IP. *J Natl Cancer Inst* **99**, 1836-44 (2007).
55. Eeles, R.A. *et al.* Identification of seven new prostate cancer susceptibility loci through a genome-wide association study. *Nat Genet* **41**, 1116-21 (2009).
56. Eeles, R.A. *et al.* Multiple newly identified loci associated with prostate cancer susceptibility. *Nat Genet* **40**, 316-21 (2008).
57. Eeles, R.A. *et al.* Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. *Nat Genet* **45**, 385-91, 391e1-2 (2013).
58. Gudmundsson, J. *et al.* Genome-wide association and replication studies identify four variants associated with prostate cancer susceptibility. *Nat Genet* **41**, 1122-6 (2009).
59. Gudmundsson, J. *et al.* Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nat Genet* **39**, 631-7 (2007).
60. Gudmundsson, J. *et al.* Common sequence variants on 2p15 and Xp11.22 confer susceptibility to prostate cancer. *Nat Genet* **40**, 281-3 (2008).
61. Gudmundsson, J. *et al.* Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. *Nat Genet* **39**, 977-83 (2007).
62. Haiman, C.A. *et al.* Genome-wide association study of prostate cancer in men of African ancestry identifies a susceptibility locus at 17q21. *Nat Genet* **43**, 570-3 (2011).
63. Kote-Jarai, Z. *et al.* Seven prostate cancer susceptibility loci identified by a multi-stage genome-wide association study. *Nat Genet* **43**, 785-91.
64. Schumacher, F.R. *et al.* Genome-wide association study identifies new prostate cancer susceptibility loci. *Hum Mol Genet* **20**, 3867-75.
65. Sun, J. *et al.* Evidence for two independent prostate cancer risk-associated loci in the HNF1B gene at 17q12. *Nat Genet* **40**, 1153-5 (2008).
66. Takata, R. *et al.* Genome-wide association study identifies five new susceptibility loci for prostate cancer in the Japanese population. *Nat Genet* **42**, 751-4 (2010).

67. Thomas, G. *et al.* Multiple loci identified in a genome-wide association study of prostate cancer. *Nat Genet* **40**, 310-5 (2008).
68. Yeager, M. *et al.* Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat Genet* **39**, 645-9 (2007).
69. Delaneau, O., Marchini, J. & Zagury, J.F. A linear complexity phasing method for thousands of genomes. *Nat Methods* **9**, 179-81 (2012).
70. Howie, B.N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* **5**, e1000529 (2009).
71. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-1 (2010).
72. Kicinski, M., Vangronsveld, J. & Nawrot, T.S. An epidemiological reappraisal of the familial aggregation of prostate cancer: a meta-analysis. *PLoS One* **6**, e27130 (2011).
73. Albright, F. *et al.* Prostate cancer risk prediction based on complete prostate cancer family history. *Prostate* **75**, 390-8 (2015).
74. Amin Al Olama, A. *et al.* Risk Analysis of Prostate Cancer in PRACTICAL, a Multinational Consortium, Using 25 Known Prostate Cancer Susceptibility Loci. *Cancer Epidemiol Biomarkers Prev* **24**, 1121-9 (2015).
75. Zhong, H. & Prentice, R.L. Bias-reduced estimators and confidence intervals for odds ratios in genome-wide association studies. *Biostatistics* **9**, 621-34 (2008).
76. Merico, D., Isserlin, R., Stueker, O., Emili, A. & Bader, G.D. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS One* **5**, e13984 (2010).
77. Joshi-Tope, G. *et al.* Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res* **33**, D428-32 (2005).
78. Schaefer, C.F. *et al.* PID: the Pathway Interaction Database. *Nucleic Acids Res* **37**, D674-9 (2009).
79. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-9 (2000).
80. Romero, P. *et al.* Computational prediction of human metabolic pathways from the complete human genome. *Genome Biol* **6**, R2 (2005).
81. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-50 (2005).
82. Kandasamy, K. *et al.* NetPath: a public resource of curated signal transduction pathways. *Genome Biol* **11**, R3 (2010).
83. Thomas, P.D. *et al.* PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res* **13**, 2129-41 (2003).
84. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**, e164 (2010).
85. Wang, L., Jia, P., Wolfinger, R.D., Chen, X. & Zhao, Z. Gene set analysis of genome-wide association studies: methodological issues and perspectives. *Genomics* **98**, 1-8 (2011).

86. Wang, K., Li, M. & Hakonarson, H. Analysing biological pathways in genome-wide association studies. *Nat Rev Genet* **11**, 843-54 (2010).
87. Wang, K., Li, M. & Bucan, M. Pathway-based approaches for analysis of genomewide association studies. *Am J Hum Genet* **81**, 1278-83 (2007).
88. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-504 (2003).